IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

JASMAN SINGH
29-11-2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- The data was analyzed using the following steps:
  - Data Collection: Used web scraping and the SpaceX API to gather information.
  - Exploratory Data Analysis (EDA): Cleaned the data, created visualizations, and explored the data interactively.
  - Machine Learning: Built models to make predictions based on the data.
  - Summary: Shared all key findings from the analysis.
- Key takeaways from the analysis:
  - Data Gathering: Collected useful data from public sources.
  - EDA Findings: Found the most important features for predicting successful launches.
  - Machine Learning Results: Identified the best model to show which factors are most important for successful launches.

Introduction

- The goal is to see if the new company Space Y can compete with SpaceX.
- Key questions to answer:
    - Cost Estimation: What is the best way to predict the total cost of launches by estimating successful rocket landings?
    - Launch Location: Where is the most suitable location for rocket launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from 2 sources:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - WebScraping(https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - The collected data was enhanced by creating a landing outcome label derived from the outcome data after summarizing and analyzing key features.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Executive Summary

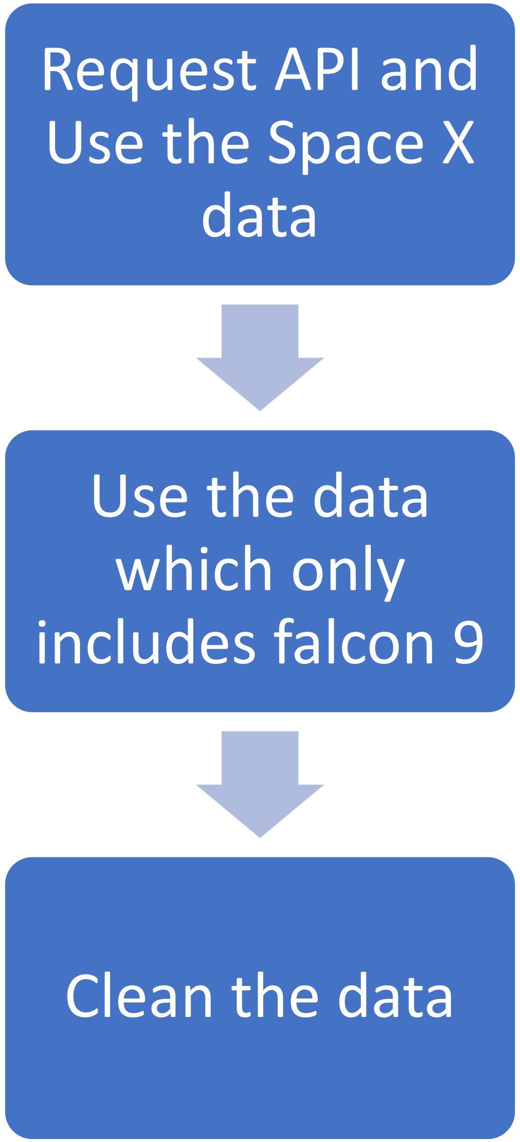
- Perform predictive analysis using classification models
  - Model Building:
    - Used machine learning algorithms like Logistic Regression, Decision Trees, KNN and Support Vector Machines to train model on the data.
  - Evaluation:
    - Assess the model's performance using metrics like accuracy.

Data Collection

- Data was collected from 2 sources:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - WebScraping(https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- You need to present your data collection process use key phrases and flowcharts

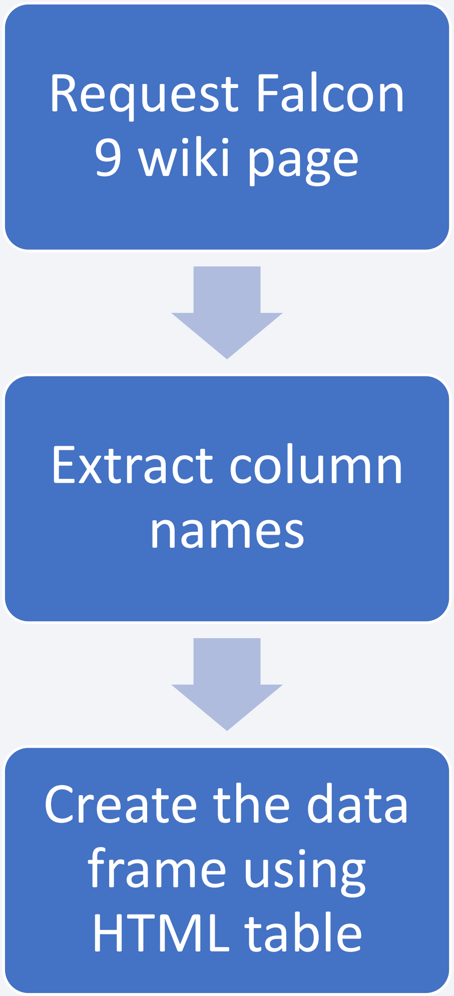
Data Collection

- SpaceX provides a public API to access data, which can then be utilized.
- This API was used following the process outlined in the flowchart and the data was saved for further use.



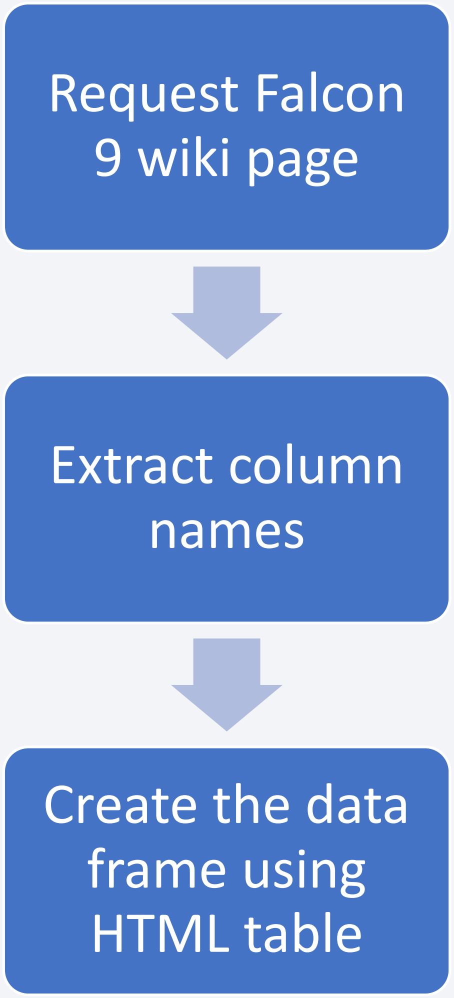
Data Collection – SpaceX API

- Data from SpaceX launches is also collected from Wikipedia;
- This is the link
- https://github.com/JasmanSingh01/Applied-Data-Science-Capstone/blob/main/capstondatascience/jupyter-labs-webscraping.ipynb



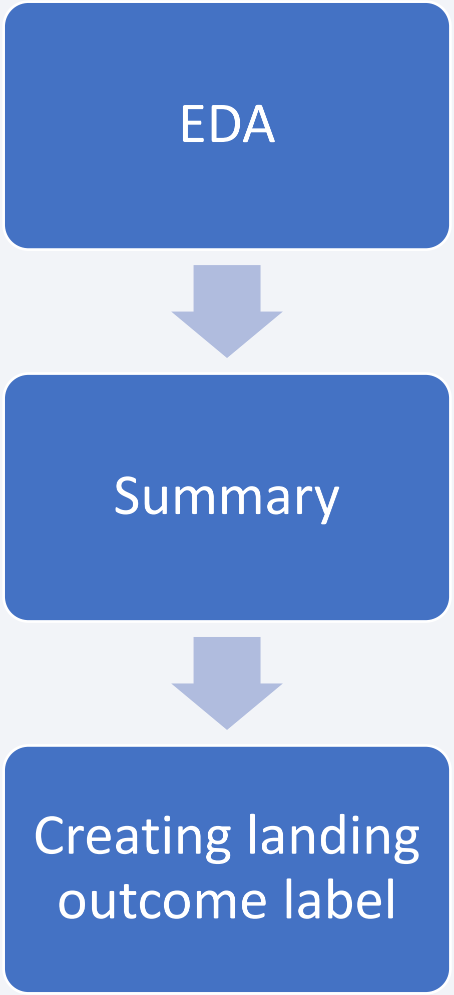
Data Collection - Scraping

- First, we requested Wikipedia page of falcon 9 using BeautifulSoup.
- Then we extract the Falcon 9 launch records HTML table from Wikipedia
- Then we parse the table and convert it into a Pandas data frame
- This is the link



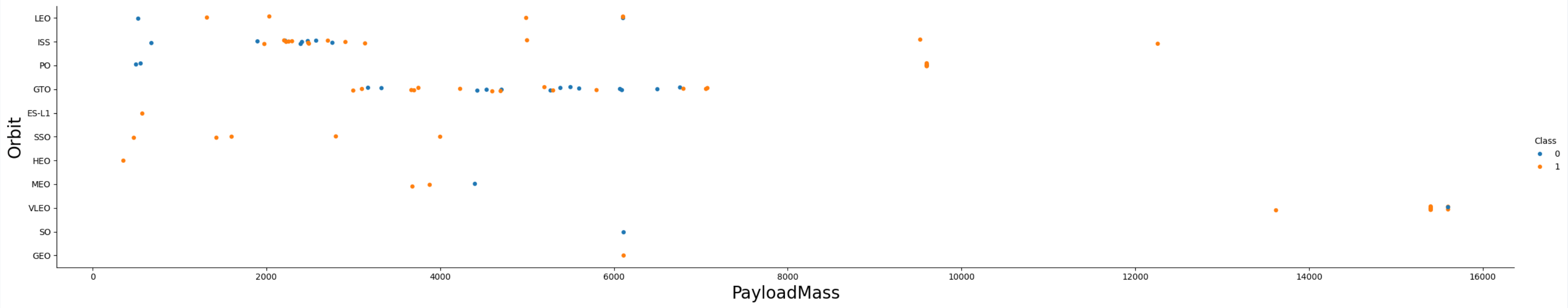
Data Wrangling

- The dataset was first explored using Exploratory Data Analysis (EDA).
- Summaries were generated for launches per site, occurrences of each orbit, and mission outcomes for each orbit type.
- A landing outcome label was then created based on the Outcome column.
- This is the link



EDA with Data Visualization

- Scatterplots and barplots were used to explore the data and visualize relationships between pairs of features:
  - Orbit vs. Flight Number
  - Launch Site vs. Flight Number
  - Launch Site vs. Payload Mass
  - Orbit vs. Success rate



- This is the Link

EDA with SQL

- The following SQL queries were performed:
  - Names of the unique launch sites in the space mission;
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Rank of the count of landing outcomes (such as Failure (drone ship) or Success (groundpad)) between the date 2010-06-04 and 2017-03-20.
- This is the Link

Build an Interactive Map with Folium

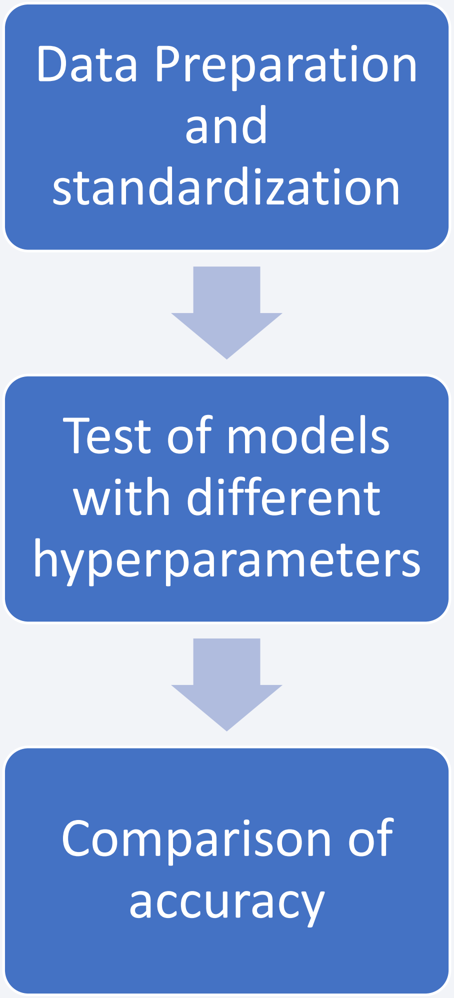
- Markers, circles, lines and marker clusters were used with Folium Maps
  - Markers indicate points like launch sites;
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center.
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site.
  - Lines are used to indicate distances between two coordinates.
- This is the Link

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.
- This is the Link

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.
- This is the Link



Results

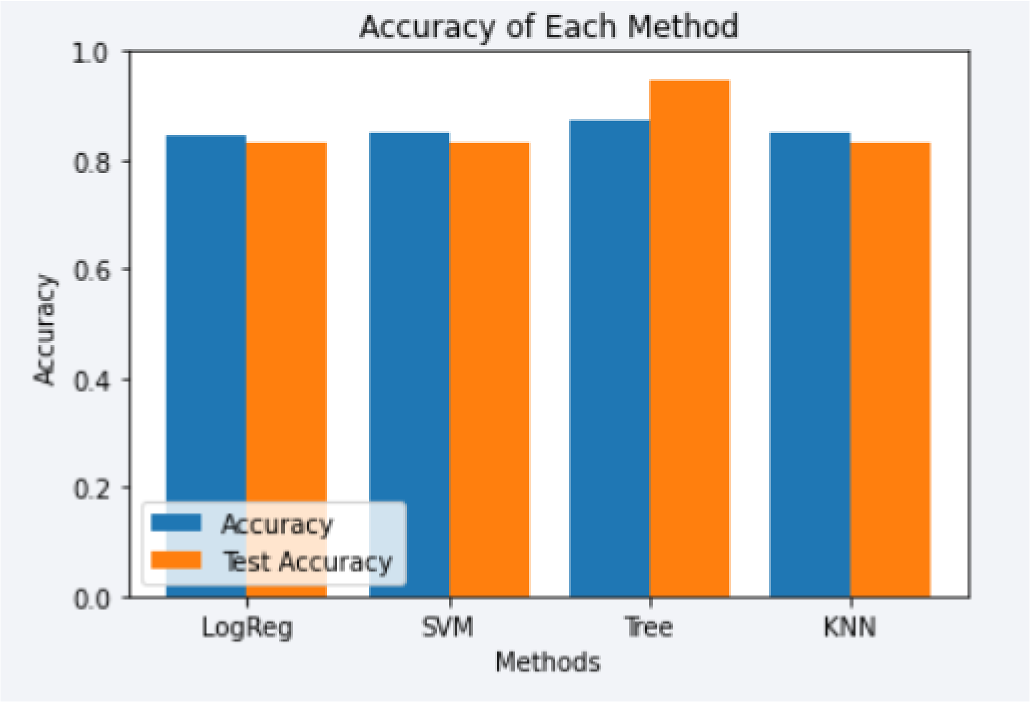
- Exploratory data analysis results:
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
  - Almost 100% of mission outcomes were successful;
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of landing outcomes became as better as years passed.

Results

- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Most launches happens at east cost launch sites.

Results

- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.

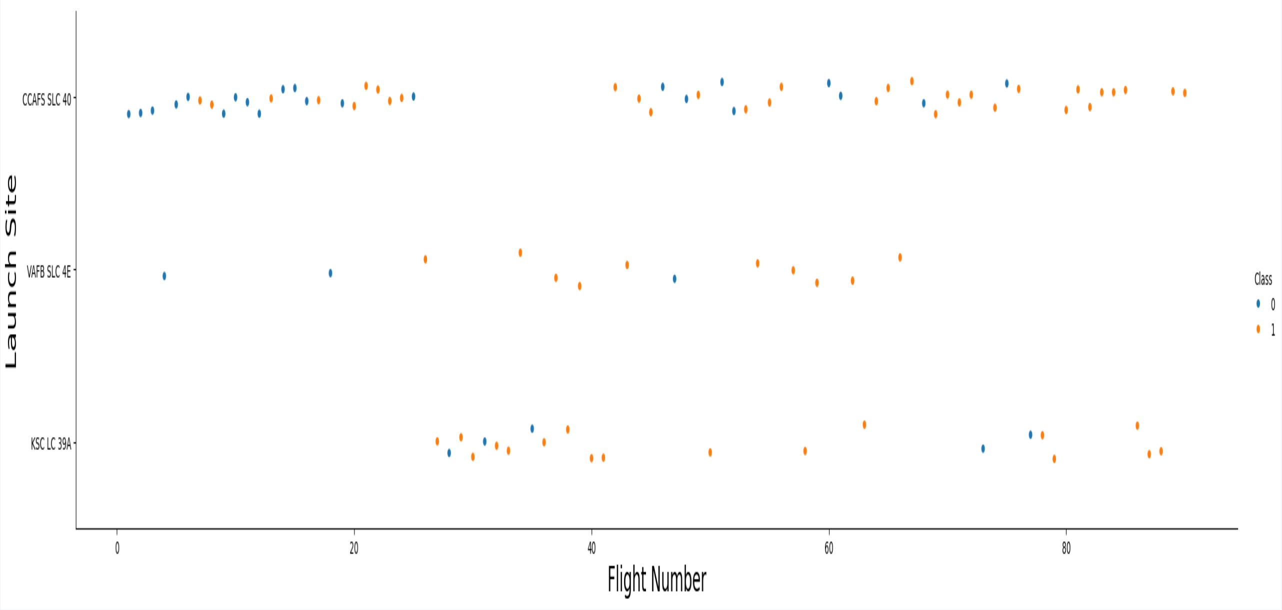


Section 2

Insights drawn from EDA

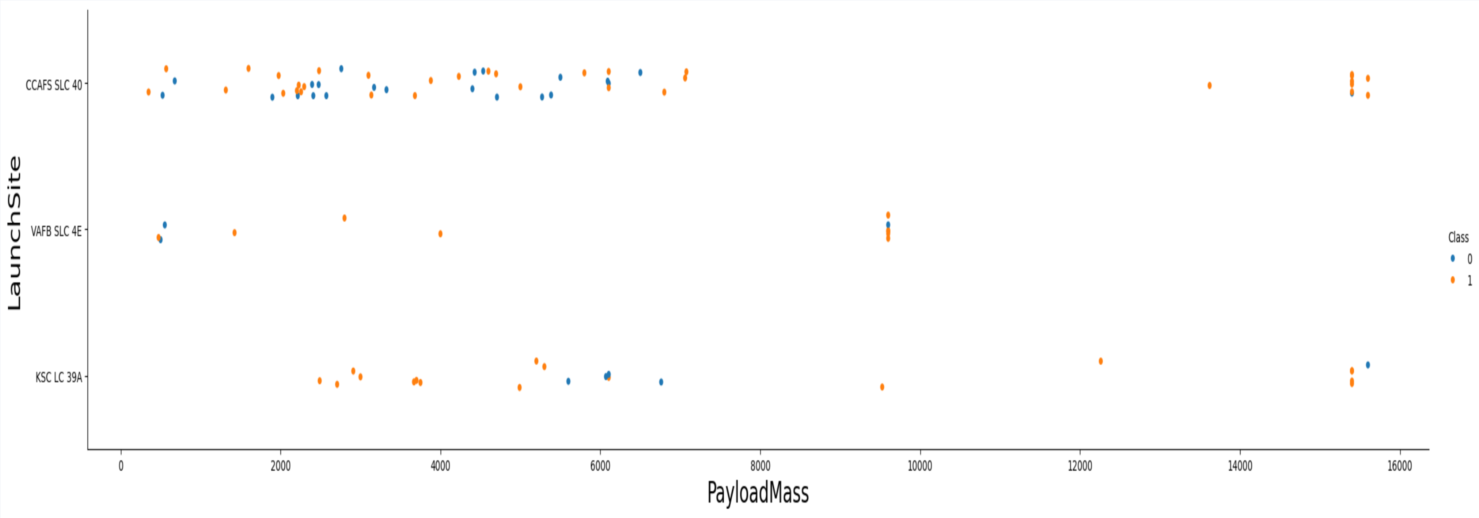
Flight Number vs. Launch Site

- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful
- In second place VAFB SLC 4E and third place KSC LC 39A
- It's also possible to see that the general success rate improved over time.



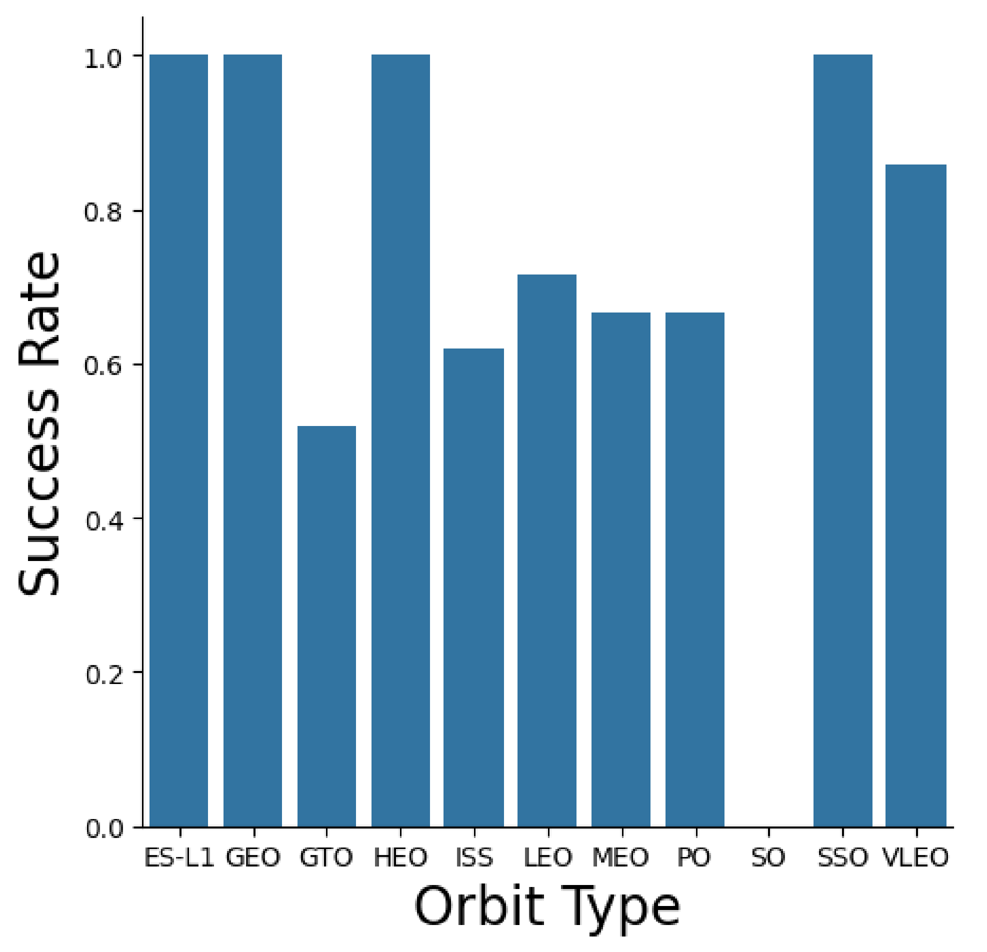
Payload vs. Launch Site

- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.



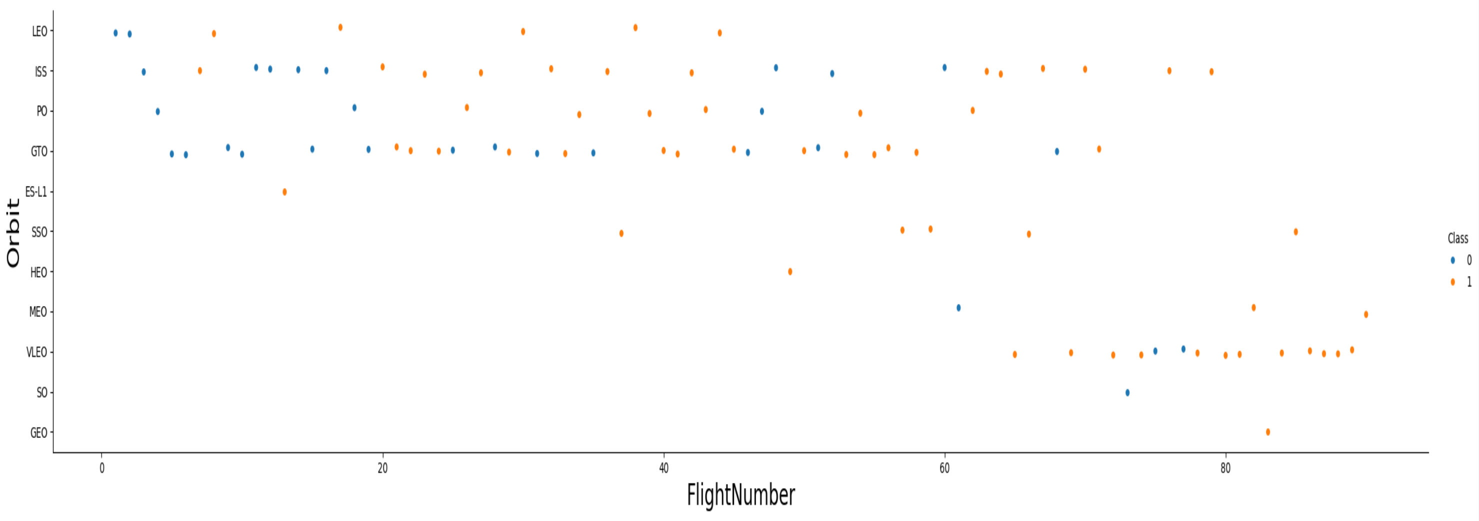
Success Rate vs. Orbit Type

- The biggest success rates happens to orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Followed by:
  - VLEO (above 80%)
  - LFO (above 70%)



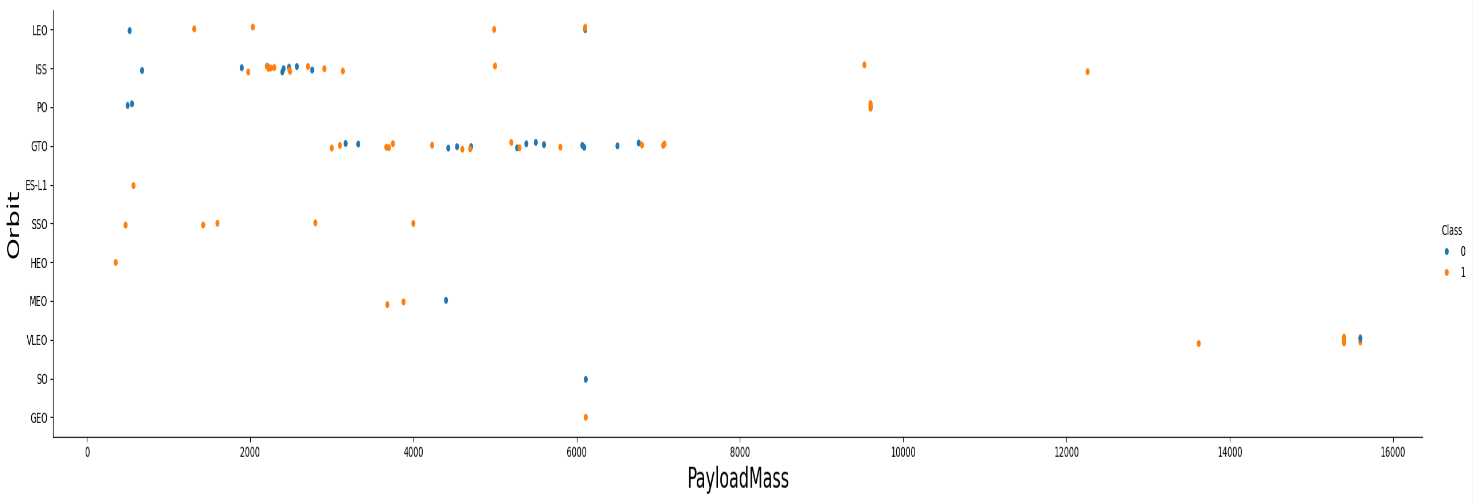
Flight Number vs. Orbit Type

- Apparently, success rate improved over time to all orbits
- VLEO orbit seems a new business opportunity, due to recent increase of its frequency.



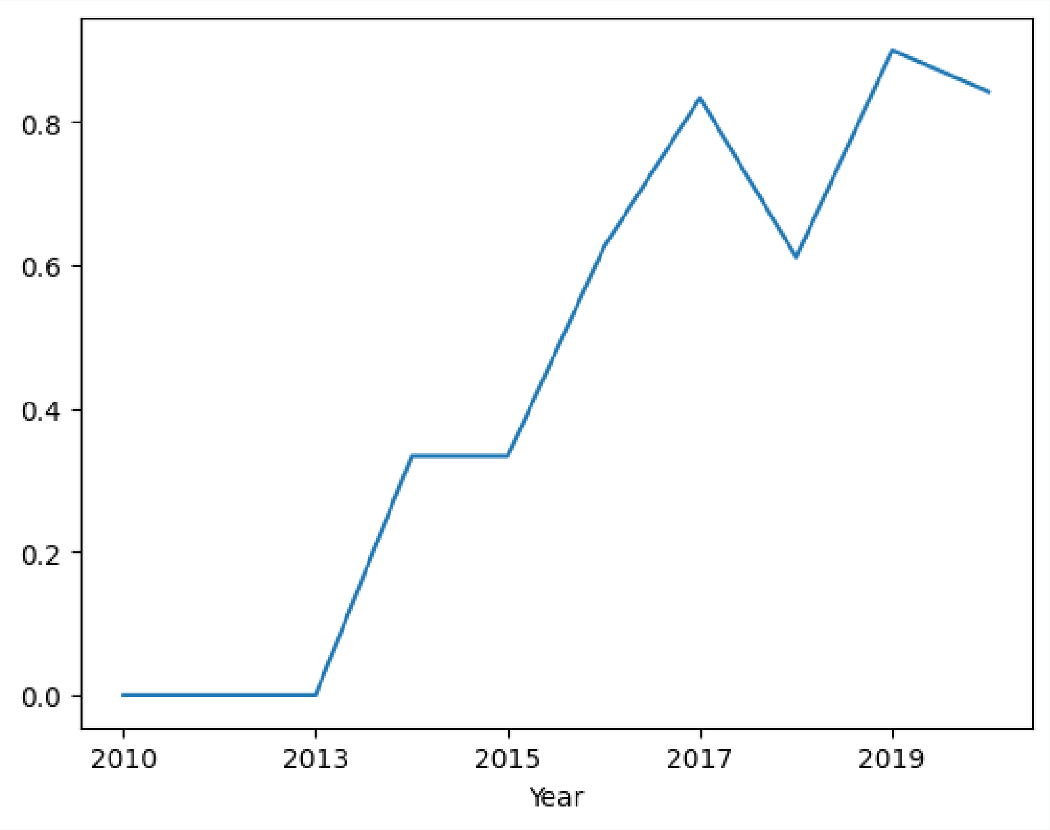
Payload vs. Orbit Type

- Apparently, there is no relation between payload and success rate to orbit GTO
- ISS orbit has the widest range of payload and a good rate of success
- There are few launches to the orbits SO and GEO



Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020
- It seems that the first three years were a period of adjusts and improvement of technology.



All Launch Site Names

- According to data, there are four launch sites:

Launch_Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

- They are obtained by selecting unique occurrences of “launch_site” values from the dataset.

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- Here we can see five samples of Cape Canaveral launches.

Total Payload Mass

- Total payload carried by boosters from NASA:

TOTAL_PAYLOAD
111268

- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:

AVG_PAYLOAD
2928.4

- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.

First Successful Ground Landing Date

- First successful landing outcome on groundpad:

FIRST_SUCCESS_GP
2015-12-22

- By filtering data by successful landing outcome on ground pad and getting the minimum value for date it's possible to identify the first occurrence, that happened on 12/22/2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- Selecting distinct booster versions according to the filters above, these 4 are the result.

Total Number of Successful and Failure Mission Outcomes

- Number of successful and failure mission outcomes:

Mission_Outcome	QTY
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass

Booster_Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- These are the boosters which have carried the maximum payload mass registered in the dataset.

2015 Launch Records

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- The list above has the only two occurrences.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Ranking of all landing outcomes between the date 2010-06-04 and 2017 03-20:

Landing_Outcome	count_outcomes
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

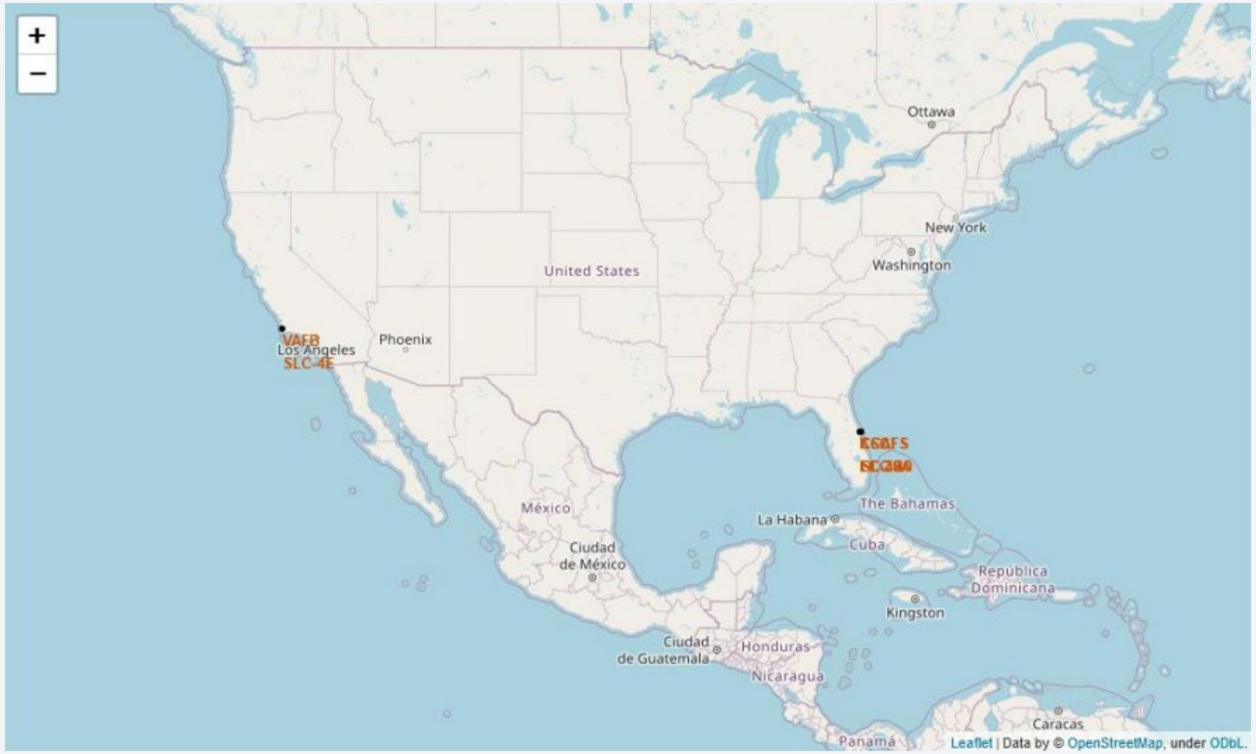
- This view of data alerts us that “No attempt” must be taken in account.

Section 3

Launch Sites Proximities Analysis

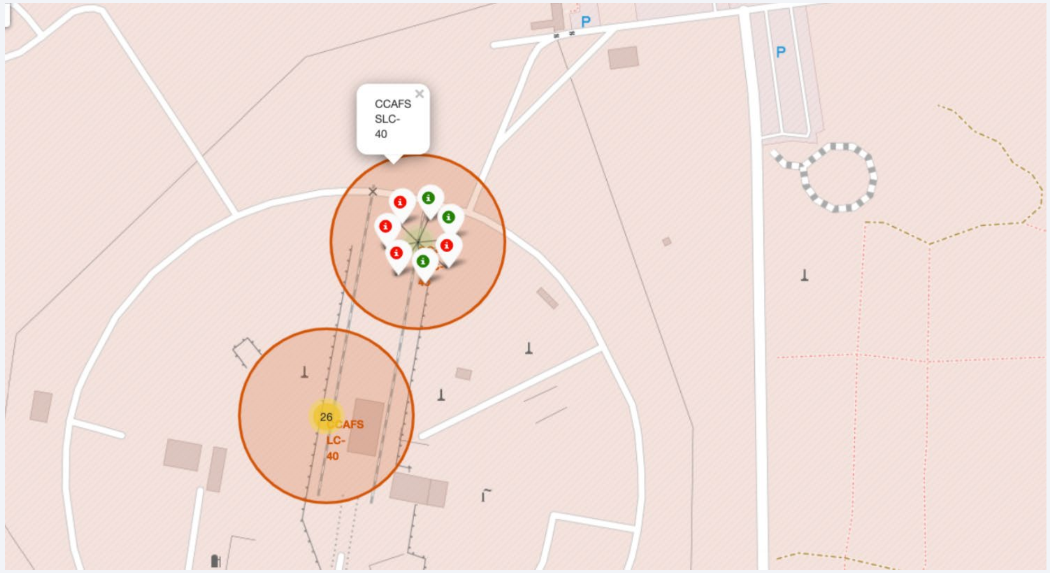
All launch sites

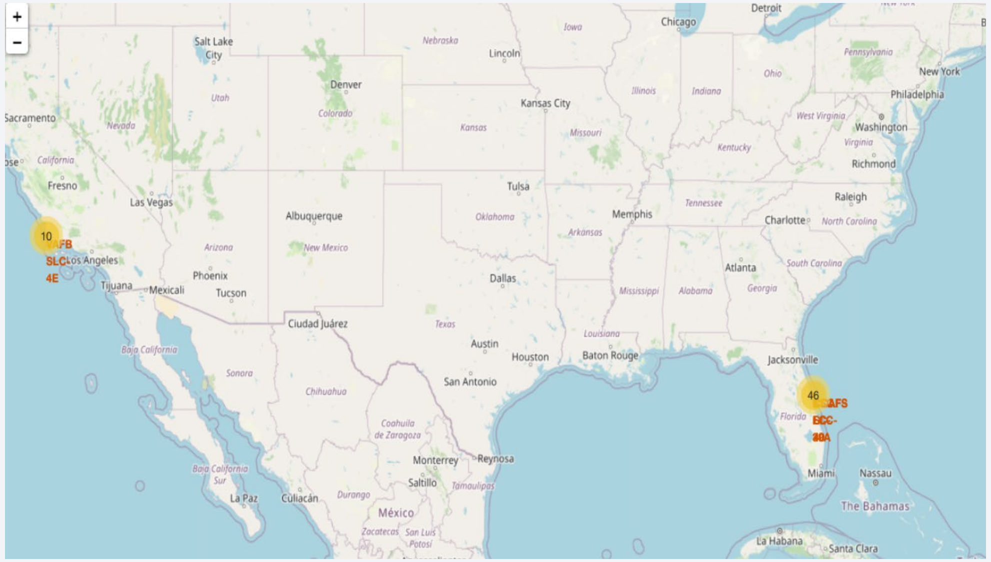
- Launch sites are near sea, probably by safety, but not too far from roads and railroads.



Launch Outcomes by Site

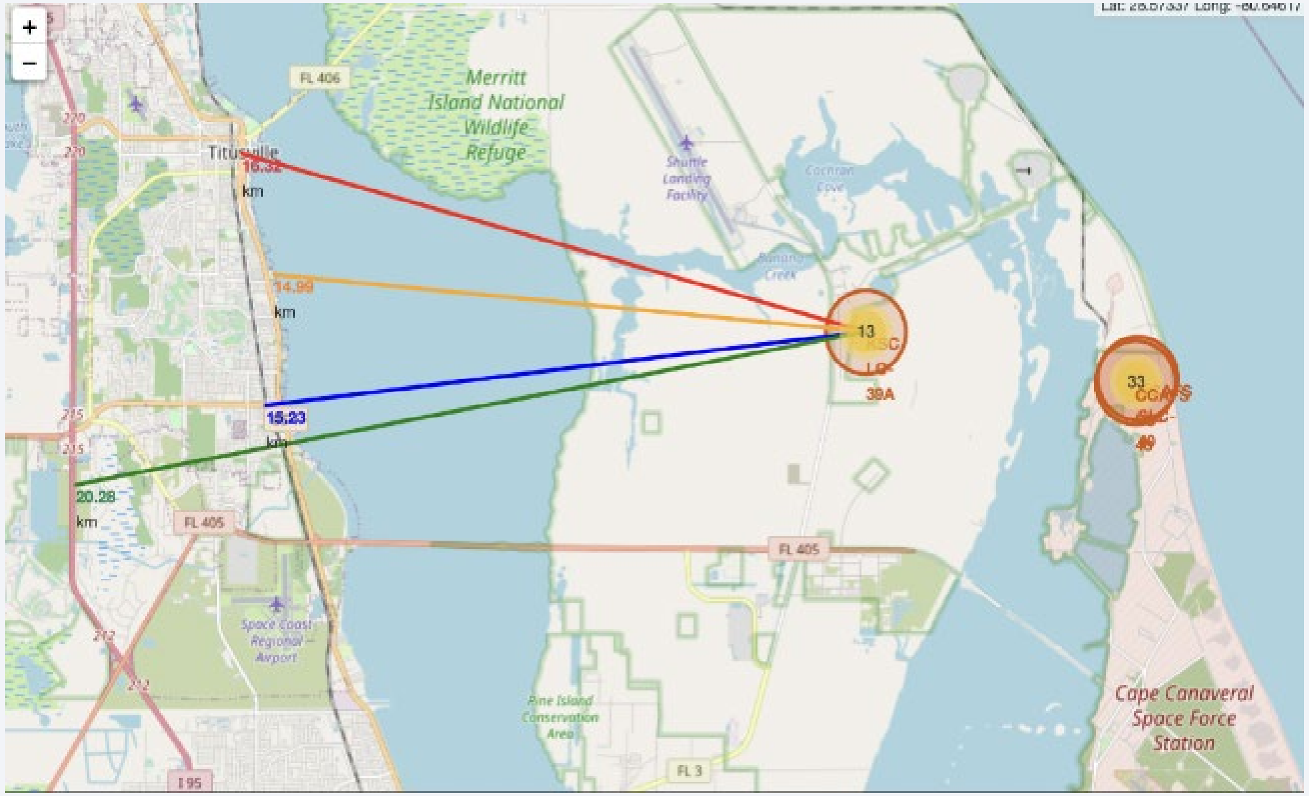
- Example of KSC LC-39A launch site launch outcomes.





- Green markers indicate successful and red ones indicate failure

Logistics and Safety



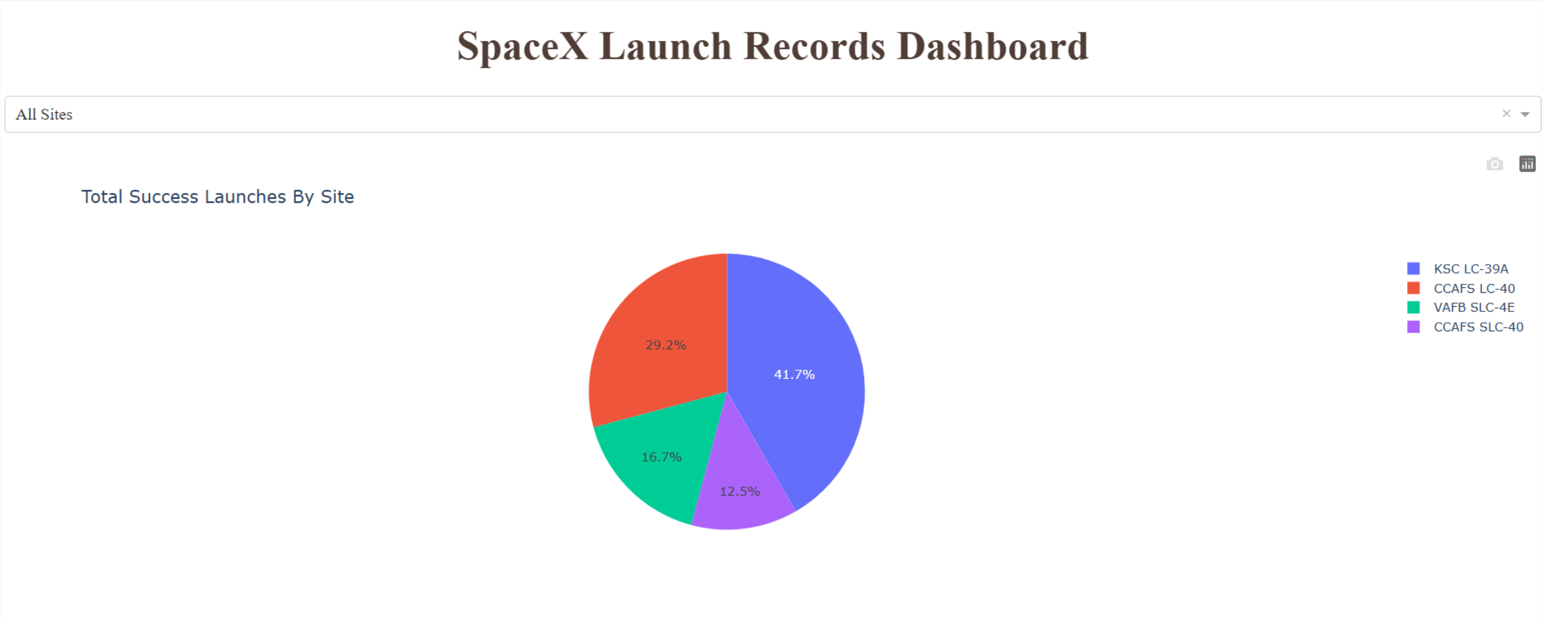
- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

Section 4

Build a Dashboard with Plotly Dash

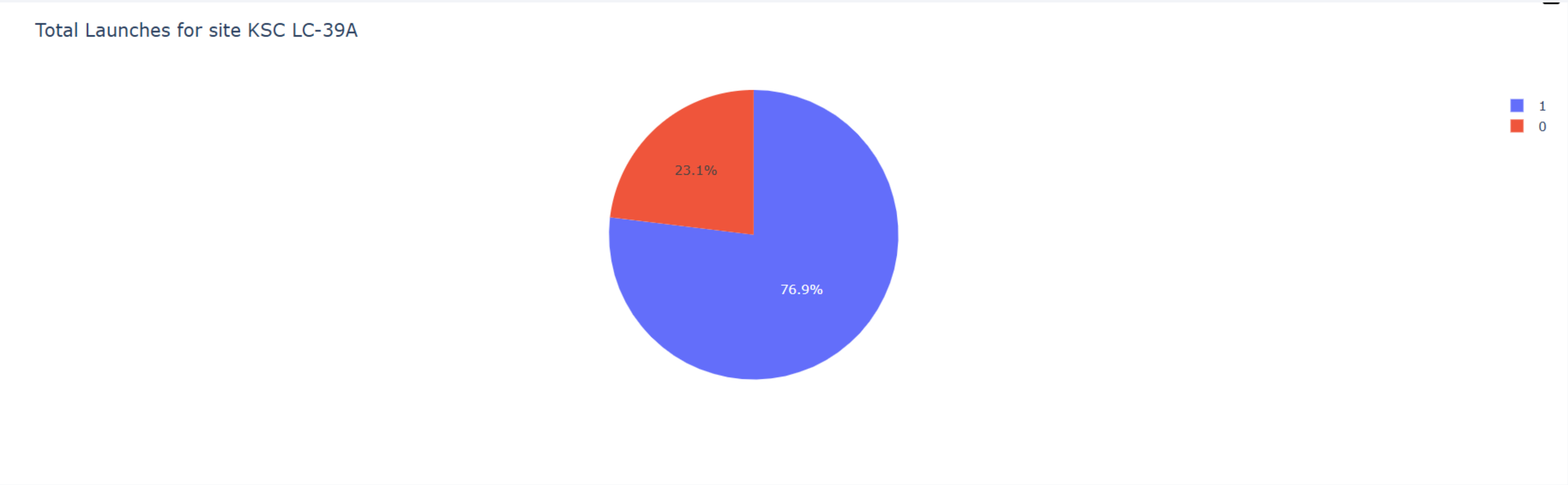
Successful Launches by Site

- The place from where launches are done seems to be a very important factor of success of missions.



Launch Success Ratio for KSC LC-39A

- 76.9% of launches are successful in this site.



Payload vs. Launch Outcome

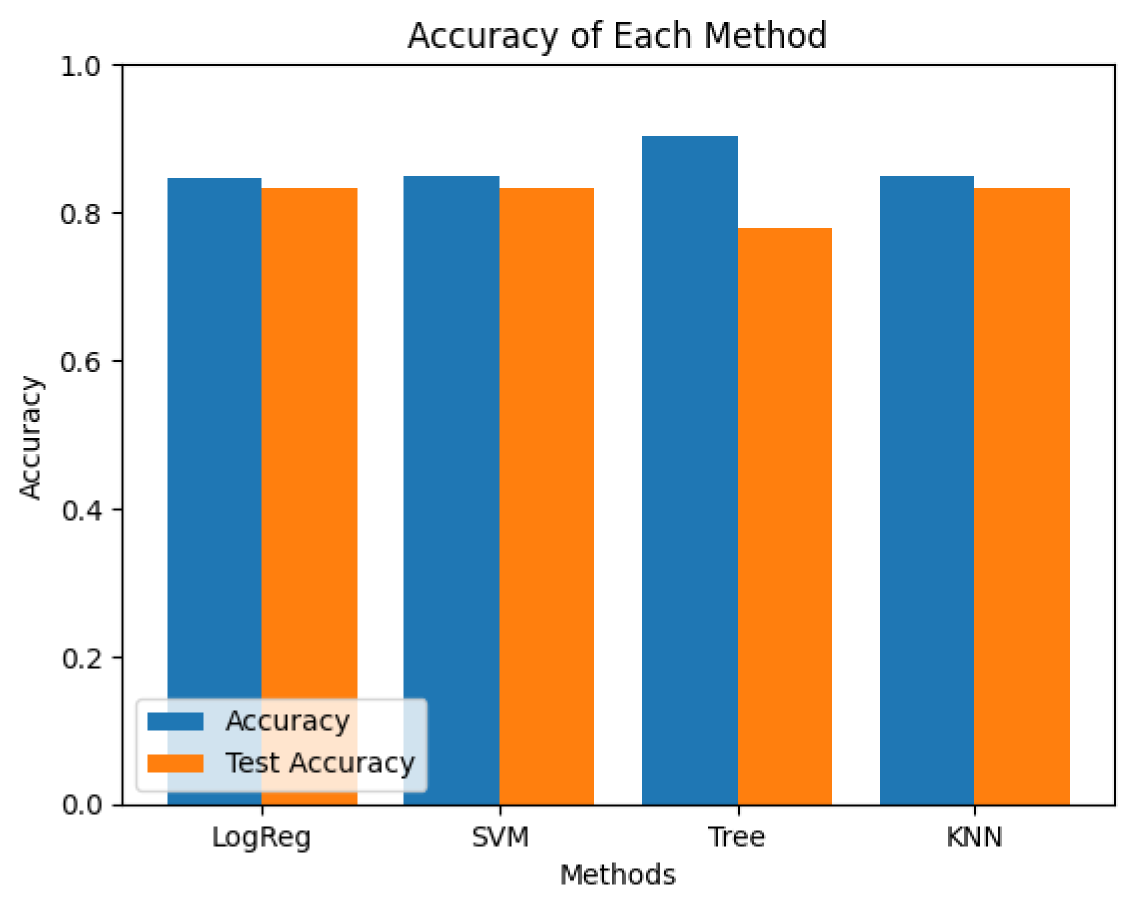
- Payloads under 6,000kg and FT boosters are the most successful combination.

Section 5

Predictive Analysis (Classification)

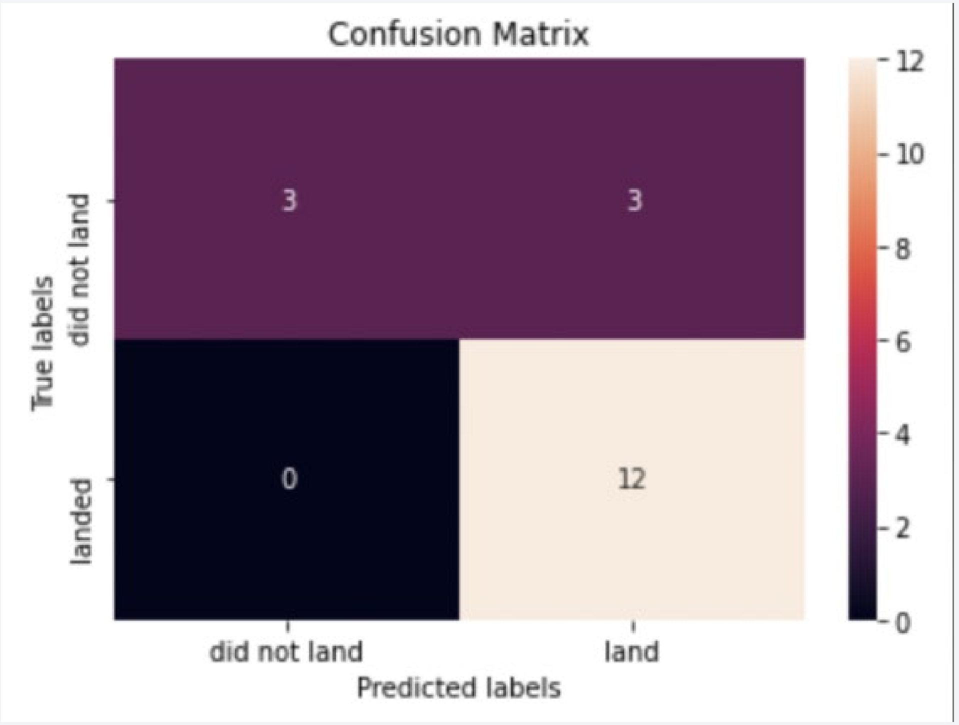
Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.



Conclusions

- Different data sources were analyzed, refining conclusions along the process
- The best launch site is KSC LC-39A
- Launches above 7,000kg are less risky
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

- This is the Link for whole repository including all csv files, codes, etc.

Thank you!